Analyzing the impact of storm events on U.S. population health and economy

Synopsis:

In examining which types of events are most harmful with respect to population health across the US, the analysis found that excessive heat, tornadoes and flash floods have had the highest number of fatalities from 1996 to 2011, and floods, excessive heat and tornadoes have had the highest number of injuries. When examining the impact per occurence of each event, excessive heat still has the highest rate of fatalities, but rip current and avalanche rise into the top three as well.

Examining which types of events have the greatest economic consequences across the US, the analysis found floods, hurricanes/typhoons and storm surges/tides have caused the greatest combined property and crop damage from 1996 to 2011. Drought meanwhile caused the most crop damage. When examining the impact per occurence of each event, floods drop down to fourth, with hurricanes/typhoons and storm surges/tides taking 1 and 2 and tropical storms joining the top three.

Data Processing

Read in data and subset it to pull out state, event type, begin date, fatalities, injuries, crop damage and property damage, to make analysis faster.

```
st <- read.csv("repdata%2Fdata%2FStormData.csv.bz2", stringsAsFactors = FALSE)
st1 <- st[, c(1, 2, 7, 8, 23:28)]
```

Convert BGN_DATE to Date class and subset data after January 1996, because the NWS didn't start recording all events until 1996.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
st1$BGN_DATE <- as.Date(st1$BGN_DATE, "%m/%d/%Y %H:%M:%S")
st1 <- st1 %>% mutate(Year = format(st1$BGN_DATE, "%Y")) %>%
        filter(Year >= 1996)
```

Do some initial quick cleaning on the EVTYPE column to get rid of summary data, make all entries upper case, and remove spaces at the beginning of the entries.

```
st1$EVTYPE <- lapply(st1$EVTYPE, function(v){
        if (is.character(v)) return(toupper(v))
        else return(v)
    })
msk <- grepl("SUMMARY", st1$EVTYPE)
st1 <- st1[!msk, ]
st1$EVTYPE <- gsub("^ ", "", st1$EVTYPE)
```

Next, subset the data further in order to be able to answer the first question: which types of events are most harmful with respect to population health? Start by finding which event types had fatalities and subset st1 for only those with more than 24 fatalities (the mean is 24.27).

```
deaths <- st1 %>% group_by(EVTYPE) %>% summarise(fatalities =
            sum(FATALITIES)) %>% filter(fatalities > 24)
q1 <- filter(st1, EVTYPE %in% deaths$EVTYPE)
```

Replace unofficial event types with the official name. The only unofficial event type unchanged is "Fog", because it could either be Dense Fog or Freezing Fog, and there isn't enough information to determine which.

```
q1$EVTYPE <- gsub("EXTREME COLD$", "EXTREME COLD/WIND CHILL", q1$EVTYPE)
q1$EVTYPE <- gsub("HEAVY SURF/HIGH SURF", "HIGH SURF", q1$EVTYPE)
q1$EVTYPE <- gsub("HURRICANE$", "HURRICANE/TYPHOON", q1$EVTYPE)
q1$EVTYPE <- gsub("RIP CURRENTS", "RIP CURRENT", q1$EVTYPE)
q1$EVTYPE <- gsub("WINTER WEATHER/MIX", "WINTER WEATHER", q1$EVTYPE)
q1$EVTYPE <- gsub("TSTM WIND", "THUNDERSTORM WIND", q1$EVTYPE)
q1$EVTYPE <- gsub("LANDSLIDE", "DEBRIS FLOW", q1$EVTYPE)
q1$EVTYPE <- gsub("URBAN/SML STREAM FLD", "FLASH FLOOD", q1$EVTYPE)
```

Then subset and process the data for question two: Which types of events have the greatest economic consequences?

First, create new columns with total propdmg and cropdmg using the exponent column. To do this, create two vectors p and c with the number corresponding to the exponent letter in PROPDMGEXP and CROPDMGEXP. Then create the new columns by multiplying the vectors by the PROPDMG and CROPDMG columns.

```
exp <- function(v) {
    B <- grepl("B", v)
    K <- grepl("K", v)
    M <- grepl("M", v)
    v[B] <- 1000000000
    v[K] <- 1000
    v[M] <- 1000000
    v[!B & !K & !M] <- 1
    as.numeric(v)
}
p <- exp(st1$PROPDMGEXP)
c <- exp(st1$CROPDMGEXP)
st1 <- mutate(st1, TotalPROPDMG = p*PROPDMG, TotalCROPDMG = c*CROPDMG)
```

Find which event types had property and crop damage and subset st1 for only those with greater than the mean property or crop damage (1.008e+09 and 9.547e+07, respectively).

```
pdamage <- st1 %>% group_by(EVTYPE) %>% summarise(propdmg =
                sum(TotalPROPDMG)) %>% filter(propdmg > 1.008e+09)
cdamage <- st1 %>% group_by(EVTYPE) %>% summarise(cropdmg =
                sum(TotalCROPDMG)) %>% filter(cropdmg > 9.547e+07)
q2 <- filter(st1, EVTYPE %in% pdamage$EVTYPE | EVTYPE %in% cdamage$EVTYPE)
```

Replace unofficial event types with the official name.

```
q2$EVTYPE <- gsub("EXTREME COLD$", "EXTREME COLD/WIND CHILL", q2$EVTYPE)
q2$EVTYPE <- gsub("^FREEZE", "FROST/FREEZE", q2$EVTYPE)
q2$EVTYPE <- gsub("HURRICANE$", "HURRICANE/TYPHOON", q2$EVTYPE)
q2$EVTYPE <- gsub("STORM SURGE$", "STORM SURGE/TIDE", q2$EVTYPE)
q2$EVTYPE <- gsub("WILD/FOREST FIRE", "WILDFIRE", q2$EVTYPE)
q2$EVTYPE <- gsub("TSTM WIND", "THUNDERSTORM WIND", q2$EVTYPE)
```

Results

Q1: Across the US, which types of events (EVTYPE) are most harmful with respect to population health?

(Perform this analysis using the q1 data set.)

Excessive Heat, Tornadoes and Flash Floods have the highest number of fatalities from 1996 to 2011.

```
q1 <- group_by(q1, EVTYPE)
fi <- q1 %>% summarise(Fatalities = sum(FATALITIES), Injuries =
            sum(INJURIES), Frequency = n()) %>% arrange(desc(Fatalities)) %>%
            rename("EventType" = EVTYPE)

head(fi, 10)
```

```
## # A tibble: 10 x 4
##                   EventType Fatalities Injuries Frequency
##                       <chr>      <dbl>    <dbl>     <int>
##  1          EXCESSIVE HEAT       1797     6391      1656
##  2                 TORNADO       1511    20667     23154
##  3             FLASH FLOOD        915     1753     54392
##  4               LIGHTNING        651     4141     13204
##  5             RIP CURRENT        542      503       734
##  6                   FLOOD        414     6758     24248
##  7       THUNDERSTORM WIND        371     5029    210071
##  8 EXTREME COLD/WIND CHILL        240      103      1619
##  9                    HEAT        237     1222       716
## 10               HIGH WIND        235     1083     19909
```

Tornadoes, Floods and Excessive Heat meanwhile have the highest number of injuries from 1996 to 2011.

```
fi <- arrange(fi, desc(Injuries))
head(fi, 10)
```

```
## # A tibble: 10 x 4
##            EventType Fatalities Injuries Frequency
##                <chr>      <dbl>    <dbl>     <int>
##  1           TORNADO       1511    20667     23154
##  2             FLOOD        414     6758     24248
##  3    EXCESSIVE HEAT       1797     6391      1656
##  4 THUNDERSTORM WIND        371     5029    210071
##  5         LIGHTNING        651     4141     13204
##  6       FLASH FLOOD        915     1753     54392
##  7 HURRICANE/TYPHOON        125     1321       258
##  8      WINTER STORM        191     1292     11317
##  9              HEAT        237     1222       716
## 10         HIGH WIND        235     1083     19909
```

When examining the impact per occurence of each event, excessive heat still has the highest rate of fatalities, at an average of 1.09 per occurence. Other events rise into the second and third positions, however, with rip current and avalanche coming in second and third with an average 0.79 and 0.59 fatalities per occurence, respectively.

```
fi2 <- fi %>% mutate(FatalityRate = Fatalities/Frequency, InjuryRate =
                        Injuries/Frequency) %>% arrange(desc(FatalityRate))
```

```
head(fi2, 10)
```

```
## # A tibble: 10 x 6
##                  EventType Fatalities Injuries Frequency FatalityRate
##                      <chr>      <dbl>    <dbl>     <int>        <dbl>
##  1                TSUNAMI         33      129        20    1.6500000
##  2          EXCESSIVE HEAT       1797     6391      1656    1.0851449
##  3             RIP CURRENT        542      503       734    0.7384196
##  4               AVALANCHE        223      156       378    0.5899471
##  5       HURRICANE/TYPHOON        125     1321       258    0.4844961
##  6                    HEAT        237     1222       716    0.3310056
##  7         COLD/WIND CHILL         95       12       539    0.1762523
##  8 EXTREME COLD/WIND CHILL        240      103      1619    0.1482397
##  9               HIGH SURF        132      198       954    0.1383648
## 10                     FOG         60      712       532    0.1127820
## # ... with 1 more variables: InjuryRate <dbl>
```

Looking at injury rates, tsunamis, hurricanes and typhoons, and excessive heat cause the most injuries per event.

```
fi2 <- arrange(fi2, desc(InjuryRate))
head(fi2, 10)
```

```
## # A tibble: 10 x 6
##            EventType Fatalities Injuries Frequency FatalityRate InjuryRate
##                <chr>      <dbl>    <dbl>     <int>        <dbl>      <dbl>
##  1           TSUNAMI         33      129        20   1.65000000  6.4500000
##  2 HURRICANE/TYPHOON        125     1321       258   0.48449612  5.1201550
##  3    EXCESSIVE HEAT       1797     6391      1656   1.08514493  3.8592995
##  4              HEAT        237     1222       716   0.33100559  1.7067039
##  5               FOG         60      712       532   0.11278195  1.3383459
##  6           TORNADO       1511    20667     23154   0.06525870  0.8925888
##  7       RIP CURRENT        542      503       734   0.73841962  0.6852861
##  8    TROPICAL STORM         57      338       682   0.08357771  0.4956012
##  9         AVALANCHE        223      156       378   0.58994709  0.4126984
## 10          WILDFIRE         75      911      2732   0.02745242  0.3334553
```

Q2: Across the US, which types of events have the greatest economic consequences?

(Perform this analysis using q2 data set.)

Floods, hurricanes/typhoons and storm surges/tides caused the greatest combined property and crop damage from 1996 to 2011.

```
damage <- q2 %>% group_by(EVTYPE) %>% summarise(PropertyDmg = sum(TotalPROPDMG),
            CropDmg = sum(TotalCROPDMG), Frequency = n()) %>%
            mutate(TotalDamage = CropDmg + PropertyDmg) %>%
            arrange(desc(TotalDamage)) %>% rename("EventType" = EVTYPE)
head(damage)
```

```
## # A tibble: 6 x 5
##           EventType  PropertyDmg    CropDmg Frequency  TotalDamage
##               <chr>        <dbl>      <dbl>     <int>        <dbl>
## 1             FLOOD 143944833550 4974778400     24248 148919611950
```

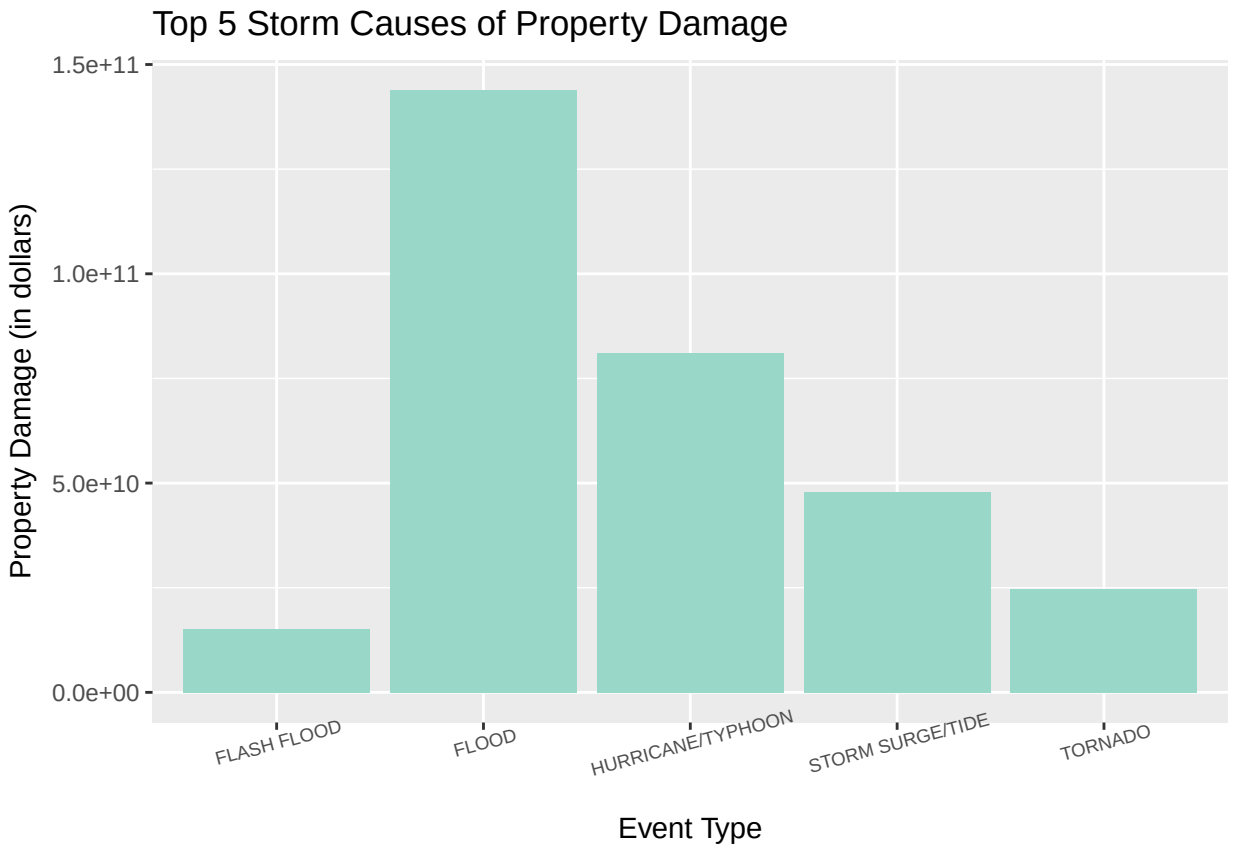
```
## 2 HURRICANE/TYPHOON  81118659010 5349282800       258  86467941810
## 3  STORM SURGE/TIDE  47834724000     855000       401  47835579000
## 4           TORNADO  24616945710  283425010     23154  24900370720
## 5              HAIL  14595143420 2476029450    207715  17071172870
## 6       FLASH FLOOD  15222253910 1334901700     51000  16557155610
```

Floods, hurricanes/typhoons and storm surges/tides have caused the greatest property damage from 1996 to 2011.

```
damageP <- arrange(damage, desc(PropertyDmg))
head(damageP, 10)
```

```
## # A tibble: 10 x 5
##            EventType  PropertyDmg    CropDmg Frequency  TotalDamage
##                <chr>        <dbl>      <dbl>     <int>        <dbl>
##  1             FLOOD 143944833550 4974778400     24248 148919611950
##  2 HURRICANE/TYPHOON  81118659010 5349282800       258  86467941810
##  3  STORM SURGE/TIDE  47834724000     855000       401  47835579000
##  4           TORNADO  24616945710  283425010     23154  24900370720
##  5       FLASH FLOOD  15222253910 1334901700     51000  16557155610
##  6              HAIL  14595143420 2476029450    207715  17071172870
##  7 THUNDERSTORM WIND   7868810880  952246350    210071   8821057230
##  8          WILDFIRE   7760449500  402255130      4175   8162704630
##  9    TROPICAL STORM   7642475550  677711000       682   8320186550
## 10          HIGH WIND   5247860360  633561300     19909   5881421660
```

```
ptop5 <- damageP[1:5, ]
library(ggplot2)
g <- ggplot(ptop5, aes(x = EventType, y = PropertyDmg))
g + geom_bar(stat = "identity", fill = "#99d8c9") + labs(x = "Event Type",
                            y = "Property Damage (in dollars)") +
    theme(axis.text.x = element_text(angle = 15, size = 7)) +
    ggtitle('Top 5 Storm Causes of Property Damage')
```

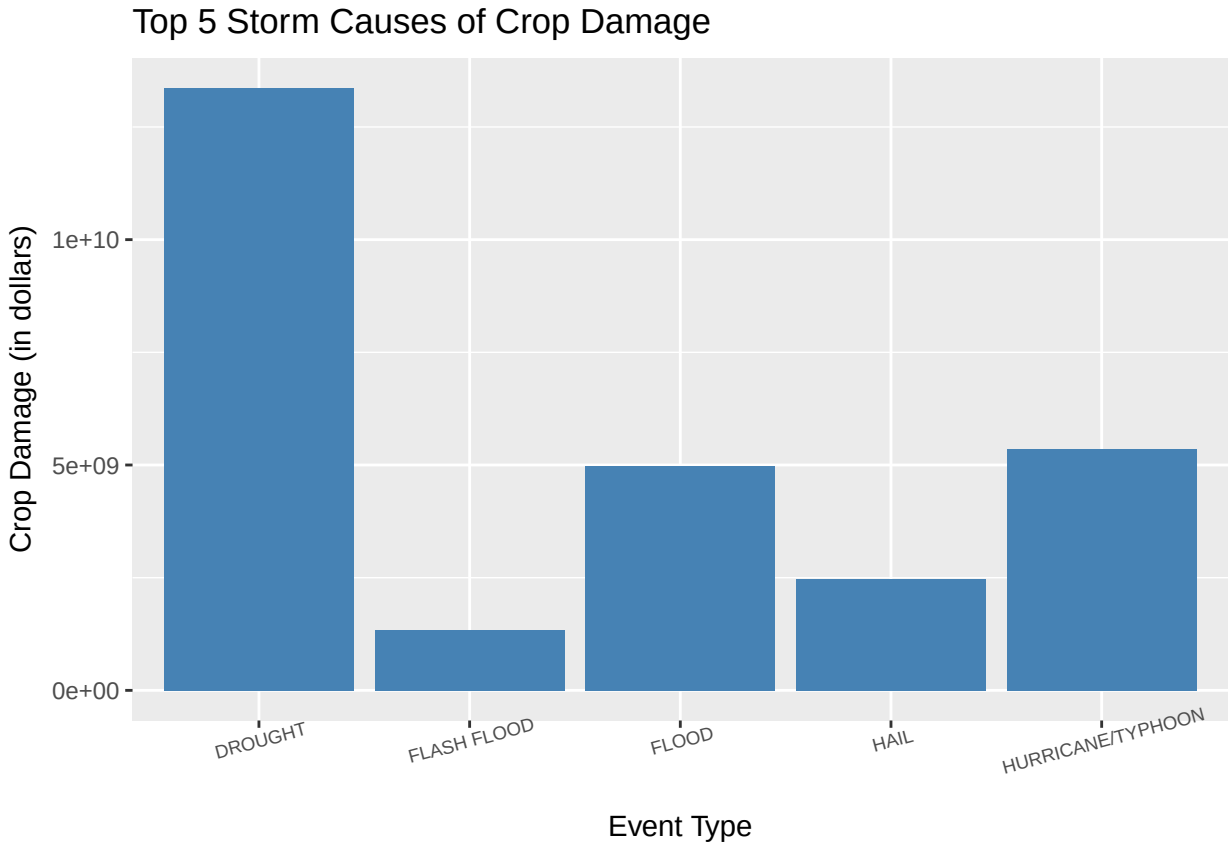


Drought has meanwhile caused the most crop damage, with hurricanes/typhoons and floods still in the top three.

```
damageC <- arrange(damage, desc(CropDmg))
head(damageC, 10)
```

```
## # A tibble: 10 x 5
##                  EventType  PropertyDmg     CropDmg Frequency  TotalDamage
##                      <chr>        <dbl>       <dbl>     <int>        <dbl>
##  1                DROUGHT   1046101000 13367566000      2433  14413667000
##  2      HURRICANE/TYPHOON  81118659010  5349282800       258  86467941810
##  3                  FLOOD 143944833550  4974778400     24248 148919611950
##  4                   HAIL  14595143420  2476029450    207715  17071172870
##  5            FLASH FLOOD  15222253910  1334901700     51000  16557155610
##  6 EXTREME COLD/WIND CHILL     19760400  1308973000       617   1328733400
##  7           FROST/FREEZE     10680000  1250911000      1412   1261591000
##  8      THUNDERSTORM WIND   7868810880   952246350    210071   8821057230
##  9             HEAVY RAIN    584864440   728169800     11528   1313034240
## 10         TROPICAL STORM   7642475550   677711000       682   8320186550
```

```
ctop5 <- damageC[1:5, ]
h <- ggplot(ctop5, aes(x = EventType, y = CropDmg))
h + geom_bar(stat = "identity", fill = "steel blue") + labs(x = "Event Type",
    y = "Crop Damage (in dollars)") +
    theme(axis.text.x = element_text(angle = 15, size = 7)) +
    ggtitle('Top 5 Storm Causes of Crop Damage')
```



When examining the economic impact per occurence of each event, floods drop down to fourth, with hurricanes/typhoons and storm surges/tides taking 1 and 2. Tropical storms rise to the third ranking.

```
damage <- damage %>% mutate(TotalDmgRate = TotalDamage/Frequency) %>%
                     arrange(desc(TotalDmgRate))
head(damage)
```

```
## # A tibble: 6 x 6
##                 EventType  PropertyDmg     CropDmg Frequency  TotalDamage
##                     <chr>        <dbl>       <dbl>     <int>        <dbl>
## 1       HURRICANE/TYPHOON  81118659010  5349282800       258  86467941810
## 2        STORM SURGE/TIDE  47834724000      855000       401  47835579000
## 3          TROPICAL STORM   7642475550   677711000       682   8320186550
## 4                   FLOOD 143944833550  4974778400     24248 148919611950
## 5                DROUGHT   1046101000 13367566000      2433  14413667000
## 6 EXTREME COLD/WIND CHILL     19760400  1308973000       617   1328733400
## # ... with 1 more variables: TotalDmgRate <dbl>
```